BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 7: e112040
doi: 10.3897/biss.7.112040

OPEN ACCESS

Conference Abstract

TaxonWorks in its 10th Year: What's new, what's next?

Deborah L Paul‡, Matthew Yoder‡

‡ University of Illinois, Champaign, United States of America

Corresponding author: Deborah L Paul (dlpaul@illinois.edu)

Received: 01 Sep 2023 | Published: 05 Sep 2023

Citation: Paul DL, Yoder M (2023) TaxonWorks in its 10th Year: What's new, what's next? Biodiversity Information Science and Standards 7: e112040. https://doi.org/10.3897/biss.7.112040

Abstract

The Species File Group (SFG) endeavors to build tools and community structures that empower researchers and collections staff in their long-term collective efforts to gather, share, and learn from biodiversity data. One such tool is TaxonWorks, now in its 10th year of development. TaxonWorks provides a collaborative workbench where scientists, collection managers, students, and volunteers capture and build on the key data and concepts we use to Describe Life (TaxonWorks motto). It provides a growing number of ways to share descriptions, from Darwin Core Archives, to NeXML-formatted observations and keys, to checklists, and bibliographies.

What's New?

We have expanded the data model of TaxonWorks, added new tools and functions, and some Companion software, that is, new stand-alone code-bases.

Two major additions, Unified Filters and Cached Maps, provide developers and users (and users who are developers) the ability to run complex queries across TaxonWorks' rich data model and to display quickly computed maps for datasets of notable size, 100K or more specimen and literature-based records. For example, Cached Maps can superimpose the asserted distribution and georeferenced literature and specimen records to create interactive searchable maps (Fig. 1).

This is an open access article distributed under the terms of the CC0 Public Domain Dedication.

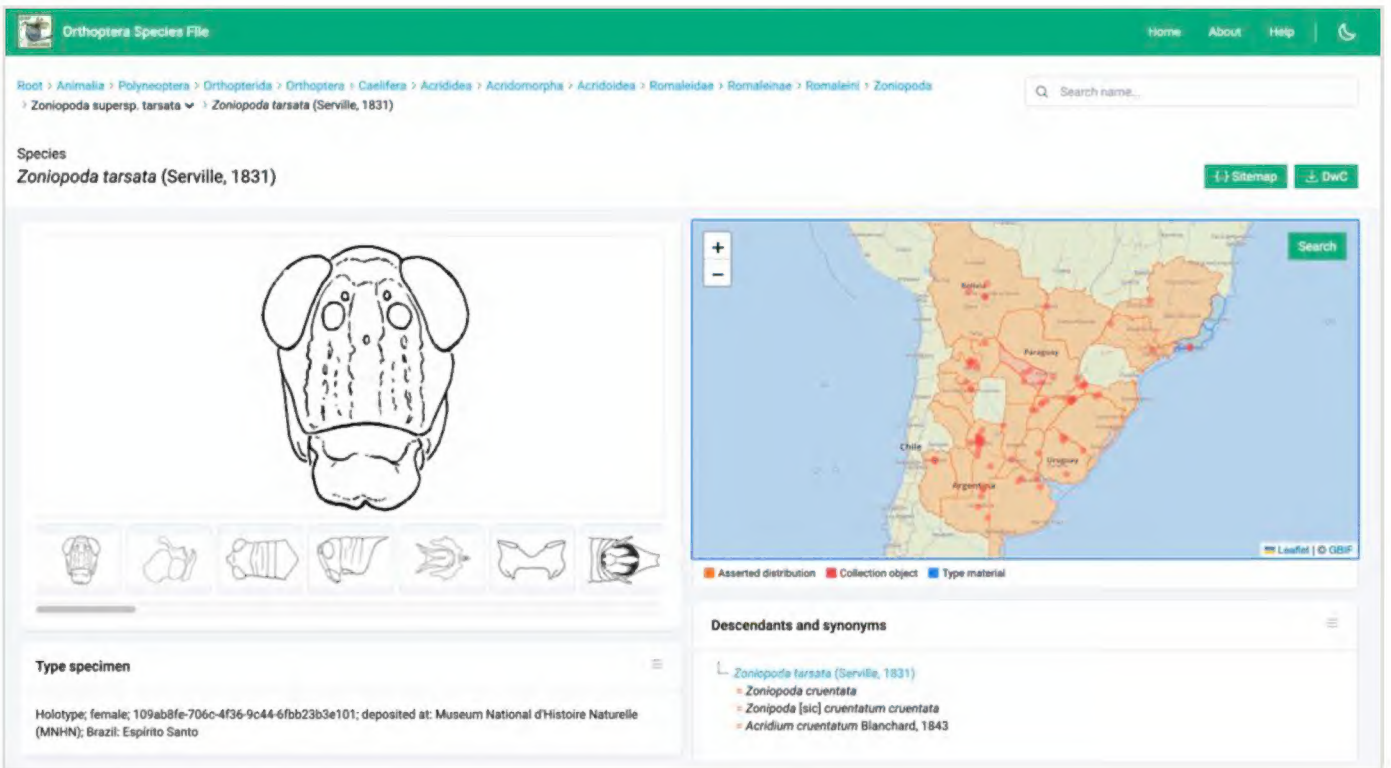

Figure 1.

Companion software TaxonPages.

In TaxonWorks, we aim to empower those working with the data with tools that help them visualize and curate information. To be able to model taxon concept relationships over time to reflect different taxonomic opinions, we added RCC-5 (Region Connection Calculus; Thau et al. 2008), which will make it possible to visualize these relationships. Similarly, we built a new visual editor (Fig. 2) for displaying, editing, and citing biological associations as recorded among specimens or taxa (or both).

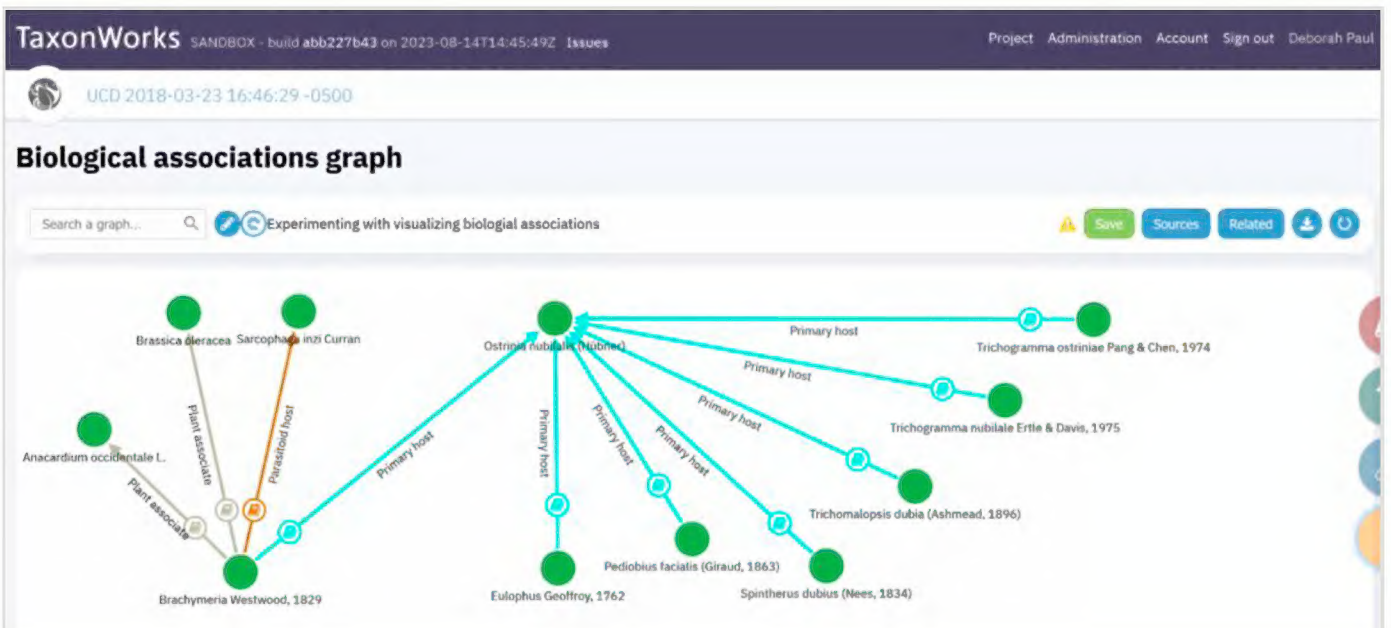

Figure 2.

Visualize, edit, cite biological associations in TaxonWorks.

Querying and enhancing data in a given database can be complex. We have worked on harmonizing the look-feel-function of the data filtering interfaces. With our Unified Filters, one can pass the results of one search to another filter (e.g., query for specimens for a

given taxonomic group and then ask for the distinct collecting events for those specimens). Then, once you filter to a given dataset, you can use our new Stepwise tasks to enhance and edit that information en-masse.

Companions code-bases extend what one can do with the data in TaxonWorks, but are also available for use with other software. For example, using our new TaxonPages code, our users can produce their own web pages for taxa (Fig. 1). TaxonPages will be used by SFG groups to make available well over 100K pages this year. They include basic Bioschema integration, links to JSON-formatted data behind every panel, and the option to download any occurrence data present, expressed as Darwin Core attributes, formatted as a CSV file. TaxonPages can be set up in minutes and served on resources like GitHub pages and our user community can customize their content.

Finally, the TaxonWorks external API has added a huge number of new parameters across multiple new conceptual endpoints.

What's Next?

With ten years of development, we see a maturing functionality surrounding the core concepts in TaxonWorks, like observations (e.g., traits, phylogenetic data), biological associations (e.g., host-parasite relationships), images, sources (citation management), specimens, collecting events, and collection management.

Currently, we are focusing on integration with other external services. We have produced multiple new API wrappers, notably Colrapi (wrapping Catalogue of Life Checklist Bank's API) and BellPepper wrapping the new Biodiversity Enhanced Location Services (BELS) Georeference API. These wrappers and ongoing integration with the Global Names Framework give our users the power to improve data quality, e.g., linking to external vocabularies, finding and updating out-of-date nomenclature, and visualizing what TaxonWorks collection object data looks like in the context of external aggregators like the Global Biodiversity Information Facility (GBIF) using our gbifference tool (as in the "GBIF difference").

The TaxonWorks community continues to grow, and therefore so does the diversity of the projects using it. Some of this diversity reflects the stage of projects: new projects need to rapidly create and stub new records, mid-life projects need to seek and add diverse data from a wide range of external resources, and mature projects need tools to identify and resolve outliers. For these data continuum scenarios, we foresee Stepwise tasks customized for managing these data maturity stage differences. Imagine capturing verbatim specimen determination data for medium-sized digitization projects and then parsing linkages to People, Times, and Taxa by the 10s, 100s, or 1000s at a time.

Some of the growing diversity behind the TaxonWorks community is a result of the end-of-life of similar tools. For example, the SFG was asked to look into moving data from Scratch pads into TaxonWorks. We are in the process of moving one Scratchpad instance and will make the scripts we used to do this publicly available for further development. In August 2023, we migrated 16 projects from legacy SFG software to TaxonWorks, bringing new

communities that can now join their expertise with others. As we move forward, we continue to work on distilling, synchronizing, and sharing our experiences and knowledge, via our community collective TaxonWorks Docs, embracing cultural change in support of the power in shared knowledge management.

Finally, TaxonWorks is committed to serving the needs of those describing species. We expect to see it produce new treatments based on extremely atomized, yet linked, data, recognizable by humans as the format serving those in the field for over 200 years. Fully formatted nomenclatural histories, descriptions, material examined sections, keys, figures and accompanying discussions based on tens of thousands of data points, all of which may be linked back to the natural history collection data that serve as their basis (or is indeed managed alongside those collections), are coming. So too are parallel tools that serve collection management needs, as the two processes are highly intertwined.

Keywords

open source software, taxonomy, collections, TaxonPages, application programming interface, RCC-5, data model, user interface, community collaboration, shared knowledge management

Presenting author

Deborah L. Paul

Presented at

TDWG 2023

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Thau D, Bowers S, Ludäscher B (2008) Merging taxonomies under RCC-5 algebraic articulations. Proceedings of the 2nd international workshop on Ontologies and information systems for the semantic web https://doi.org/10.1145/1458484.1458492